

EFFICIENT CONFORMAL BINARY CLASSIFICATION UNDER NEAREST NEIGHBOR

MAXWELL LOVIG

Communicated by Paul Eloe

ABSTRACT. There are many types of statistical inferences that can be used today: Frequentist, Bayesian, Fiducial, and others. However, Vovk introduced a new version of statistical inference known as Conformal Predictions. Conformal Predictions were designed to reduce the assumptions of standard prediction methods. Instead of assuming all observations are drawn independently and identically distributed, we instead assume exchangeability. Meaning, all $N!$ possible orderings of our N observations are equally likely. This is more applicable to fields such as machine learning where assumptions may not be easily satisfied. In the case of binary classification, Vovk provided the nearest neighbors (NN) measure which is a ratio of in-class versus out-of-class distance. Later on, Papodopolous introduced normalizing constants for NN for the regression case, we extend this work to the classification case. We provide an asymptotic guarantee which shows what is known empirically. The normalization of NN produces smaller confidence sets on average compared to standard NN. A small synthetic simulation is also presented to shown the viability in a non-asymptotic case.

KEYWORDS: *Efficiency, conformal predictions, set prediction, asymptotic, confidence sets*

MSC (2010): Primary 62G20

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LOUISIANA, LAFAYETTE, LA, USA
E-mail address: maxwelllovig@gmail.com

Received November 8, 2021; revised January 11, 2022.